International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IMPACT OF SUPERVISED MACHINE LEARNING FRAMEWORKS TOWARDS PROTEIN-PROTEIN INTERACTION PREDICTION

DASARADHA RAMAYYA LANKA
RESEARCH SCHOLAR DEPARTMENT OF TECHNOLOGY & COMPUTER SCIENCE, THE GLOCAL UNIVERSITY SAHARANPUR. UP.

DR. PRATAP SINGH PATWAL
PROFESSOR, DEPARTMENT IN TECHNOLOGY & COMPUTER SCIENCE, THE GLOCAL UNIVERSITY SAHARANPUR.UP.

ABSTRACT

Transfer learning techniques leverage pre-trained models on large-scale protein datasets to initialize neural network architectures for PPI prediction tasks, improving performance and generalization capabilities. Graph neural networks are tailored for modeling complex interactions in PPI networks, capturing topological properties and structural characteristics for more accurate predictions.

Furthermore, the integration of multi-omics data, including genomics, transcriptomics, and proteomics, provides a holistic view of cellular processes, augmenting the predictive accuracy of models. Efforts to improve interpretability and explainability enable researchers to understand the underlying biological mechanisms driving predicted interactions, facilitating the translation of computational predictions into actionable insights.

Standardized benchmarks and evaluation protocols ensure consistent assessment of PPI prediction methods, driving innovation and facilitating comparisons across different approaches. By integrating unsupervised and supervised machine learning frameworks, researchers can develop more robust and accurate models for predicting protein-protein interactions, contributing to a deeper understanding of cellular processes and disease mechanisms.

KEYWORDS: Supervised Machine, Learning Frameworks, Protein-Protein Interaction Prediction

INTRODUCTION

Protein-protein interaction (PPI) prediction serves as a cornerstone in bioinformatics and computational biology, offering a plethora of applications across various domains. One of the primary uses of PPI prediction lies in advancing our understanding of biological systems by elucidating the intricate networks of interactions between proteins. These interactions govern numerous cellular processes, including signal transduction, gene regulation, metabolic pathways, and immune responses. By predicting protein interactions, researchers can unravel the underlying mechanisms driving these biological processes, identify key players in cellular pathways, and gain insights into the dynamics and regulation of protein complexes. Moreover, PPI prediction facilitates the exploration of protein interaction networks at a genome-wide scale, enabling the systematic analysis of protein function, evolution, and disease mechanisms.

In addition to advancing our understanding of biological systems, PPI prediction plays a pivotal role in drug discovery and development by identifying novel drug targets and therapeutic candidates. Proteins involved in disease pathways often interact with each other to regulate cellular functions or mediate pathogenic processes. By predicting protein interactions associated with disease-related pathways or complexes, researchers can pinpoint potential drug targets for intervention. Furthermore, PPI prediction enables the identification of protein-protein interfaces and binding sites, facilitating the design and optimization of small-molecule inhibitors, peptides, or biologics that disrupt or modulate protein interactions. These computational approaches complement experimental screening efforts and accelerate the discovery of candidate drugs for treating various diseases, including cancer, infectious diseases, neurodegenerative disorders, and autoimmune conditions.

Moreover, PPI prediction serves as a valuable tool in systems biology and network medicine for modeling and simulating complex biological systems. Protein interaction networks represent the structural and functional organization of cellular processes, enabling the integration of diverse data types, such as gene expression, protein localization, and phenotype data, into unified network representations. By incorporating predicted protein interactions into network models, researchers can simulate the behavior of biological systems, predict cellular responses to perturbations, and uncover emergent properties of complex networks. Furthermore, PPI prediction facilitates the identification of functional modules, pathways, and regulatory circuits within protein interaction networks, providing insights into the modular organization and hierarchical structure of biological systems. These network-based approaches offer a holistic view of cellular processes and enable the elucidation of genotype-phenotype relationships, disease mechanisms, and drug responses in complex biological systems.

Furthermore, PPI prediction plays a crucial role in personalized medicine and precision oncology by identifying patient-specific biomarkers and therapeutic targets. Cancer is characterized by aberrant signaling pathways and dysregulated protein interactions that drive tumor initiation, progression, and metastasis. By analyzing protein interaction networks in cancer cells or patient samples, researchers can identify key driver genes, oncogenic signaling pathways, and vulnerabilities associated with specific cancer subtypes or patient populations. Moreover, PPI prediction facilitates the identification of biomarkers predictive of disease progression, treatment response, and patient outcomes, enabling personalized treatment strategies tailored to individual patients' molecular profiles. Additionally, PPI prediction aids in the discovery of synthetic lethal interactions, where the simultaneous inhibition of two interacting proteins selectively kills cancer cells with specific genetic mutations while sparing normal cells, offering promising opportunities for developing targeted cancer therapies.

Furthermore, PPI prediction serves as a valuable resource for functional annotation and prioritization of genes and proteins in genomic and proteomic studies. With the exponential growth of genomic and proteomic data, computational methods are indispensable for interpreting the functional significance of genes and proteins and prioritizing candidates for further experimental validation. By predicting protein interactions and functional associations, researchers can annotate the biological roles and functions of uncharacterized genes and proteins, infer functional modules and pathways, and prioritize candidate genes or proteins for functional studies. Additionally, PPI prediction aids in the interpretation of genome-wide association studies (GWAS) and identification of disease-associated genes and variants by integrating PPI networks with genetic and phenotypic data. These integrative approaches provide insights into the molecular mechanisms underlying complex traits and diseases and facilitate the discovery of potential therapeutic targets and biomarkers for precision medicine.

Moreover, PPI prediction has applications beyond biology and healthcare, spanning fields such as biotechnology, agriculture, and environmental science. In biotechnology, PPI prediction facilitates the design and engineering of protein complexes and molecular machines for biocatalysis, biofuel production, and pharmaceutical manufacturing. By predicting protein interactions and interfaces, researchers can engineer protein complexes with desired functions, optimize enzyme-substrate interactions, and enhance the efficiency and specificity of biotechnological processes. In agriculture, PPI prediction aids in the characterization of plant-pathogen interactions, symbiotic relationships, and stress responses, enabling the development of crops with improved yield, resilience, and nutritional value. Additionally, PPI prediction contributes to environmental science by elucidating microbial interactions, biogeochemical cycles, and ecosystem dynamics, facilitating the design of bioremediation strategies, microbial consortia, and synthetic ecosystems for environmental remediation and sustainability.

FEATURE ENGINEERING AND REPRESENTATION LEARNING:

Effective feature engineering is crucial for developing accurate PPI prediction models. Integrating unsupervised techniques for feature extraction contributes to a more comprehensive representation of protein characteristics. Moreover, recent advances in representation learning, such as word embeddings and graph embeddings, have demonstrated significant improvements in capturing the inherent relationships within biological data. These embeddings can be leveraged to enhance the feature space used by supervised models, thereby improving their ability to discern complex patterns in PPI networks.

Feature engineering and representation learning are two fundamental aspects of machine learning that play a critical role in extracting meaningful information from raw data, enabling the development of robust and accurate predictive models across various domains. Feature engineering involves the process of selecting, transforming, and creating new features from the raw data to improve the performance of machine learning algorithms. On the other hand, representation learning focuses on automatically learning feature representations directly from the data, often through deep learning architectures, without the need for manual feature engineering. Both feature engineering and representation learning techniques are essential for effectively capturing the underlying patterns and structures within data, enabling the development of models that can make accurate predictions or decisions.

Feature engineering encompasses a wide range of techniques aimed at transforming raw data into a format that is more suitable for machine learning algorithms. One common approach in feature engineering involves selecting relevant features that are most informative for the task at hand while discarding irrelevant or redundant features that may introduce noise or increase computational complexity. This process often involves domain knowledge and expertise to identify features that capture the essential characteristics of the data and are predictive of the target variable. For example, in natural language processing tasks, features such as word frequency, syntactic structure, and semantic similarity are commonly used to represent text data, while in image recognition tasks, features such as pixel intensity, color histograms, and texture descriptors may be used to represent image data.

In addition to feature selection, feature engineering also involves transforming raw features into a more meaningful representation through techniques such as scaling, normalization, and encoding. Scaling and normalization techniques ensure that features are on a similar scale, preventing features with larger magnitudes from dominating the learning process and improving the convergence of optimization algorithms. Encoding techniques, such as one-hot encoding and label encoding, are used to represent categorical variables as numerical values that can be processed by machine learning algorithms. Furthermore, feature engineering may also involve creating new features through mathematical transformations, interactions, or aggregations of existing features to capture higher-order relationships or patterns within the data.

While traditional feature engineering techniques have been widely used and effective in many applications, they often rely on manual intervention and domain expertise, which can be time-consuming and may not always capture the full complexity of the data. Representation learning, on the other hand, offers a more automated and data-driven approach to feature extraction by learning feature representations directly from the data. Representation learning algorithms, particularly deep learning architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), can automatically learn hierarchical representations of the input data through multiple layers of nonlinear transformations.

One of the key advantages of representation learning is its ability to capture complex and abstract features from raw data, enabling the development of models that can effectively generalize to unseen data and tasks. For example, in image recognition tasks, CNNs can learn hierarchical representations of image features, starting from low-level features such as edges and textures and progressing to higher-level features such as object shapes and structures. Similarly, in natural language processing tasks, RNNs can learn distributed representations of word sequences, capturing syntactic and semantic relationships between words in a sentence.

Furthermore, representation learning techniques such as word embeddings and autoencoders have been widely used to learn dense, low-dimensional representations of high-dimensional data, facilitating more efficient and effective processing by machine learning algorithms. Word embeddings, for example, represent words as dense vectors in a continuous vector space, capturing semantic similarities and relationships between words. Autoencoders, on the other hand, learn to reconstruct the input data from a compressed representation, effectively capturing the most salient features of the data while discarding noise and irrelevant information.

Despite the advantages of representation learning, it is not without its challenges. One of the main challenges is the need for large amounts of labeled data to train deep learning models effectively. Deep learning architectures, particularly deep neural networks, typically require large datasets to learn complex feature representations and avoid overfitting. However, labeled data is often scarce or expensive to obtain, particularly in domains where manual annotation is required. As a result, representation learning techniques may not always be feasible or practical for all applications, especially in domains with limited labeled data.

Another challenge of representation learning is the interpretability of learned representations. While deep learning models are capable of learning highly

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

discriminative features from raw data, understanding and interpreting these learned representations can be challenging. Unlike traditional feature engineering techniques, which often result in interpretable and human-readable features, deep learning models learn complex and abstract representations that may be difficult to interpret or understand. As a result, the black-box nature of deep learning models can hinder their adoption in domains where interpretability and transparency are essential, such as healthcare and finance.

Despite these challenges, representation learning techniques continue to advance rapidly, driven by innovations in deep learning architectures, optimization algorithms, and computational resources. Recent developments such as self-supervised learning, generative adversarial networks (GANs), and transformer architectures have demonstrated significant improvements in feature learning and representation capabilities, enabling the development of more powerful and generalizable models across a wide range of domains. Furthermore, techniques such as transfer learning and domain adaptation have been proposed to address the challenges of labeled data scarcity and domain shift, enabling the transfer of knowledge learned from one task or domain to another.

NETWORK-BASED APPROACHES:

Network-based methods have gained prominence in PPI prediction, leveraging graph theory to represent and analyze protein interactions. Unsupervised network clustering identifies densely connected subgraphs, suggesting potential protein complexes or pathways. Supervised approaches, on the other hand, can utilize graph-based features to predict interactions between proteins. Integrating both unsupervised and supervised network-based techniques allows for a more comprehensive understanding of the underlying biological systems, capturing both local and global interactions within protein networks.

Network-based approaches represent a powerful paradigm in various fields, including biology, social science, computer science, and beyond. These approaches leverage the representation of entities and their interactions as networks or graphs, where nodes represent entities, and edges represent relationships or interactions between them. Network-based approaches provide a versatile framework for analyzing complex systems, enabling the extraction of valuable insights into the structure, function, and dynamics of interconnected entities. By leveraging network representations, researchers can uncover hidden patterns, identify influential nodes, detect communities, and predict interactions, facilitating advancements in diverse domains.

In the realm of biology, network-based approaches have revolutionized our understanding of complex biological systems, including cellular processes, protein-protein interactions (PPIs), gene regulatory networks, and disease mechanisms. Biological systems are inherently networked, with molecules such as proteins, genes, and metabolites interacting with each other to carry out essential functions. Network-based approaches enable the integration of heterogeneous biological data sources, including genomic, transcriptomic, proteomic, and metabolomic data, into unified network representations, facilitating the exploration of complex relationships and interactions within biological systems. For example, protein interaction networks capture physical interactions between proteins, providing insights into the functional organization of cellular processes and the mechanisms underlying diseases. By analyzing the topology of protein interaction networks, researchers can identify critical proteins, pathways, and modules that play key roles in cellular functions and disease pathways, paving the way for the discovery of novel drug targets and therapeutic interventions.

Furthermore, network-based approaches have been instrumental in elucidating the genetic basis of complex diseases and understanding the interplay between genes, proteins, and phenotypes. Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with various diseases and traits, but interpreting the functional implications of these variants remains challenging. Network-based approaches offer a powerful framework for prioritizing candidate disease genes and variants by integrating GWAS data with protein interaction networks, gene expression data, and functional annotations. By mapping disease-associated genes onto protein interaction networks, researchers can identify disease modules and pathways enriched for genetic variants, providing insights into the molecular mechanisms underlying diseases and suggesting potential therapeutic targets. Moreover, network-based approaches enable the prediction of novel disease-associated genes and interactions through guilt-by-association principles, where genes with similar network properties or functional annotations to known disease genes are prioritized as candidate disease genes.

CHALLENGES AND FUTURE DIRECTIONS:

Despite the progress made in integrating unsupervised and supervised machine learning frameworks for PPI prediction, challenges persist. The inherent complexity and dynamic nature of biological systems pose difficulties in accurately modeling protein interactions. Additionally, the scarcity of high-quality labeled datasets limits the training of robust supervised models. Future research should focus on addressing these challenges, exploring advanced machine learning techniques, and leveraging emerging technologies like deep learning to further enhance the predictive capabilities of PPI models.

The integration of unsupervised and supervised machine learning frameworks represents a promising avenue for advancing protein-protein interaction prediction. By combining the strengths of unsupervised methods in extracting meaningful patterns from biological data with the precision of supervised approaches in learning from labeled datasets, researchers can develop more accurate and reliable models. As technology continues to evolve and biological data becomes more abundant, the synergy between unsupervised and supervised machine learning is poised to play a pivotal role in unraveling the intricacies of protein interactions and contributing to our broader understanding of cellular processes.

Challenges and future directions in various fields, including but not limited to technology, science, healthcare, and society, present intricate problems that demand innovative solutions to navigate a rapidly evolving landscape. In the realm of technology, the relentless pursuit of advancements has led to a plethora of challenges, including cybersecurity threats, data privacy concerns, and ethical dilemmas surrounding artificial intelligence (AI) and automation. As technology becomes increasingly integrated into every aspect of our lives, ensuring the security and privacy of data, mitigating the risks associated with emerging technologies, and addressing the ethical implications of AI and automation remain paramount. Moreover, the rapid pace of technological innovation necessitates continuous adaptation and upskilling of the workforce to remain competitive in the digital economy, highlighting the importance of lifelong learning and education.

In the field of science, challenges abound in addressing pressing global issues such as climate change, biodiversity loss, and infectious diseases. Climate change poses existential threats to ecosystems, economies, and societies worldwide, necessitating urgent action to mitigate greenhouse gas emissions, transition to renewable energy sources, and build resilience to climate impacts. Biodiversity loss, driven by habitat destruction, pollution, and climate change, jeopardizes ecosystem services essential for human well-being, necessitating conservation efforts, sustainable land management practices, and global cooperation to protect and restore biodiversity. Furthermore, infectious diseases, such as the COVID-19 pandemic, highlight the interconnectedness of global health security and the need for robust healthcare systems, pandemic preparedness, and equitable access to vaccines and treatments to address emerging infectious threats effectively.

In the realm of healthcare, challenges abound in providing accessible, affordable, and high-quality healthcare to all, particularly in the face of demographic changes, rising healthcare costs, and disparities in healthcare access and outcomes. Aging populations and the burden of chronic diseases pose significant challenges to healthcare systems worldwide, necessitating innovative approaches to promote healthy aging, prevent disease, and manage chronic conditions effectively. Moreover, the COVID-19 pandemic has exposed weaknesses in healthcare infrastructure, supply chains, and response capabilities, highlighting the need for resilient healthcare systems, pandemic preparedness, and investments in public health infrastructure and workforce capacity. Addressing healthcare disparities, improving access to healthcare services, and advancing health equity are critical to ensuring that all individuals have the opportunity to lead healthy and fulfilling lives.

TYPE OF PROTEIN-PROTEIN INTERACTION PREDICTION

Protein-protein interaction (PPI) prediction encompasses various computational methodologies aimed at elucidating the interactions between proteins within biological systems. These predictions serve as invaluable tools in understanding cellular processes, identifying novel drug targets, and unraveling disease mechanisms. Among the diverse range of methods employed for PPI prediction, several key types stand out, each with distinct approaches and applications. Sequence-based prediction methods constitute one significant category, leveraging information encoded in amino acid sequences to infer potential interactions between proteins. These methods typically involve the extraction of features from protein sequences, such as amino acid composition, physicochemical properties, and evolutionary conservation, followed by the training of machine learning models, such as support vector machines (SVMs) or neural networks, to classify interacting and non-interacting protein pairs. Sequence-based prediction approaches are advantageous due to their simplicity, scalability, and applicability to a wide range of proteins, making them suitable for large-scale PPI prediction tasks. However, they may struggle to capture complex interaction patterns or structural information that influence protein interactions, limiting their predictive accuracy in certain cases.

In contrast, structure-based prediction methods exploit information derived from protein structures to predict protein-protein interactions. These methods typically involve the modeling of protein structures using techniques such as homology modeling, molecular docking, or protein threading, followed by the analysis of intermolecular contacts, binding interfaces, and energy calculations to predict potential protein complexes and interactions. Structure-based prediction approaches offer the advantage of capturing detailed structural information and spatial constraints that govern protein interactions, enabling the prediction of binding affinities, interface residues, and complex stoichiometry. However, they often require accurate protein structure predictions, which may be challenging for proteins with no experimentally determined structures or those exhibiting conformational flexibility or disorder.

Furthermore, network-based prediction methods leverage information from protein interaction networks to predict protein-protein interactions. These methods exploit the topology, connectivity, and community structure of protein interaction networks to infer potential interactions between proteins based on their network properties and local neighborhood. Network-based prediction approaches often involve graph-based algorithms, such as random walk, label propagation, or network embedding techniques, to predict missing edges or interactions in protein interaction networks. Additionally, machine learning models, such as graph neural networks or deep learning architectures, can be trained on network-based features to predict protein interactions and identify functionally related protein modules or complexes. Network-based prediction methods offer the advantage of capturing global network properties and context-dependent relationships between proteins, enabling the prediction of indirect or context-specific interactions that may not be apparent from sequence or structure alone. However, they may suffer from limitations in data availability, network incompleteness, and biases in network construction, which can affect the reliability and generalization performance of prediction models.

Moreover, integrative prediction methods combine multiple data modalities, including sequence, structure, expression, and functional annotations, to improve the accuracy and coverage of PPI predictions. These methods leverage complementary information from diverse data sources to generate consensus predictions that are more robust and reliable than individual predictions from single data types. Integrative prediction approaches often involve the integration of heterogeneous data sources using machine learning algorithms, statistical models, or network-based approaches to generate integrated feature representations or consensus scores for predicting protein-protein interactions. Additionally, meta-learning techniques, such as ensemble methods or transfer learning, can be employed to combine predictions from multiple models or data sources to improve prediction performance further. Integrative prediction methods offer the advantage of leveraging complementary information from diverse data sources, enhancing the accuracy and reliability of PPI predictions. However, they may require sophisticated data integration and model fusion techniques, as well as comprehensive benchmarking and validation, to ensure the robustness and generalization performance of integrated prediction models.

CONCLUSION

The discovery of PPIs plays a vital role in Bioinformatics and Biomedical Science. Through PPIs, researchers infer the cellular mechanisms of proteins. Exploration of novel PPIs facilitates the drug discovery to identify therapeutic agents for various human ailments. Only a tiny fraction of the expected number of PPI concerning the entire protein corpus of a species is experimentally validated, that hinders the ad- vancement of respective downstream applications in Bioinformatics. Therefore, com- putational approaches complement biological experiments for alleviating deficiencies in the PPI prediction. These methodologies require different kinds of features, that play an indispensable role in deploying a machine learning framework for PPI pre- diction. Due to the increasing richness of information, GO is one of the most favored feature sources for various bioinformatic analyses, including PPI prediction.

REFERENCES

1. Taha, Kamal. (2023). Employing Machine Learning Techniques to Detect Protein-Protein Interaction: A Survey, Experimental, and Comparative Evaluations. 10.1101/2023.08.22.554321.

2. Zhang, Mengying & Su, Qiang & Lu, Yi & Zhao, Manman & Niu, Bing. (2017). Application of Machine Learning Approaches for Protein-protein Interactions Prediction. Medicinal chemistry (Shariqah (United Arab Emirates)). 13. 10.2174/1573406413666170522150940.

3. Sarkar, Debasree & Saha, Sudipto. (2019). Machine-learning techniques for the prediction of protein–protein interactions. Journal of Biosciences. 44. 10.1007/s12038-019-9909-z.

4. Khandelwal, Monika & Rout, Ranjeet & Umer, Saiyed. (2022). Protein-protein interaction prediction from primary sequences using supervised machine learning algorithm. 10.1109/Confluence52989.2022.9734190.

5. Lee, Minhyeok. (2023). Recent Advances in Deep Learning for Protein-Protein Interaction Analysis: A Comprehensive Review. Molecules. 28. 5169. 10.3390/molecules28135169.

6. Li, Shiwei & Wu, Sanan & Wang, Lin & Li, Fenglei & Jiang, H. & Bai, Fang. (2022). Recent advances in predicting protein–protein interactions with the aid of artificial intelligence algorithms. Current Opinion in Structural Biology. 73. 102344. 10.1016/j.sbi.2022.102344.

7. Lei, Haijun & Wen, Yuting & Elazab, Ahmed & Tan, Ee-Leng & Zhao, Yujia & Lei, Baiying. (2018). Protein-Protein Interactions Prediction via Multimodal Deep

Polynomial Network and Regularized Extreme Learning Machine. IEEE Journal of Biomedical and Health Informatics. PP. 1-1. 10.1109/JBHI.2018.2845866.

8. Chakraborty, Arijit & Mitra, Sajal & De, Debashis & Pal, Anindya & Ghaemi, Ferial & Ahmadian, Ali & Ferrara, Massimiliano. (2021). Determining Protein–Protein Interaction Using Support Vector Machine: A Review. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3051006.

9. Song, Bosheng & Luo, Xiaoyan & Luo, Xiaoli & Liu, Yuansheng & Niu, Zhangming & Zeng, Xiangxiang. (2022). Learning spatial structures of proteins improves protein-protein interaction prediction. Briefings in bioinformatics. 23. 10.1093/bib/bbab558.

10. Du, Tianchuan & Liao, Li & Wu, Cathy. (2016). Enhancing interacting residue prediction with integrated contact matrix prediction in protein-protein interaction. EURASIP Journal on Bioinformatics and Systems Biology. 2016. 10.1186/s13637-016-0051-z.

11. Cheng, Jianlin & Tegge, Allison & Baldi, Pierre. (2008). Machine Learning Methods for Protein Structure Prediction. Biomedical Engineering, IEEE Reviews in. 1. 41 - 49. 10.1109/RBME.2008.2008239.

12. Zhang, Long & Yu, Guo-Xian & Xia, Dawen & Wang, Jun. (2018). Protein-Protein Interactions Prediction based on Ensemble Deep Neural Networks. Neurocomputing. 324. 10.1016/j.neucom.2018.02.097.

13. Zhang, Li & Li, Wenhao & Guan, Haotian & He, Zhiquan & Cheng, Mingjun & Wang, Han. (2023). MCPI: Integrating Multimodal Data for Enhanced Prediction of Compound Protein Interactions.

14. Wang, Yan-Bin & You, Zhu-Hong & Li, Xiao & Jiang, Tong-Hai & Zhou, Xi & Wang, Lei. (2017). Predicting protein–protein interactions from protein sequences by a stacked sparse autoencoder deep neural network. Mol. BioSyst..

13. 1336-1344. 10.1039/C7MB00188F.

15. Alquran, Hiam & Al-Fahoum, Amjed & Zyout, Alaa & Qasmieh, Isam. (2023). A comprehensive framework for advanced protein classification and function prediction using synergistic approaches: Integrating bispectral analysis, machine learning, and deep learning. PLOS ONE. 18. e0295805. 10.1371/journal.pone.0295805.

16. Hu, Xiaotian & Feng, Cong & Ling, Tianyi & Chen, Ming. (2022). Deep learning frameworks for protein-protein interaction prediction. Computational and structural biotechnology journal. 20. 3223-3233. 10.1016/j.csbj.2022.06.025.

17. Albu, Alexandra-Ioana. (2012). An Approach for Predicting Protein-Protein Interactions using Supervised Autoencoders. Procedia Computer Science. 207. 2023-2032. 10.1016/j.procs.2022.09.261.

18. Jamasb, Arian & Day, Ben & Cangea, Catalina & Lio, Pietro & Blundell, Tom. (2011). Deep for Protein–Protein Interaction Site Prediction. 10.1007/978-1-0716-1641-3_16.

19. Sun, Tanlin & Zhou, Bo & Lai, Luhua & Pei, Jianfeng. (2017). Sequence-based prediction of protein protein interaction using a deep-learning algorithm. BMC Bioinformatics. 18. 10.1186/s12859-017-1700-2.